

Adrien Francais<sup>1</sup>, Aurélien Vesin<sup>2</sup>,  
Jean-François Timsit<sup>3</sup>

## How to conduct clinical research studies using high quality-clinical databases in the in critical care

*Como realizar estudos de pesquisa clínica em terapia intensiva utilizando base de dados de alta qualidade*

1. INSERM U823/ Team 11 (Outcome of Cancer and Critical Illnesses); Coordinating Center of the OUTCOMEREA database, Albert Bonniot Institute, La Tronche, France.  
2. INSERM U823/ Team 11 (Outcome of Cancer and Critical Illnesses); Coordinating Center of the OUTCOMEREA database, Albert Bonniot Institute, La Tronche, France.  
3. INSERM U823/ Team 11 (Outcome of Cancer and Critical Illnesses); Coordinating Center of the OUTCOMEREA database, Albert Bonniot Institute, La Tronche, France and Medical Intensive Care Unit, University Hospital Albert Michallon, Grenoble, France.

Received from the Albert Bonniot Institute, La Tronche, France.

Submitted on May 16, 2008  
Accepted on July 10, 2008

### Address for correspondence:

Prof Jean-François Timsit (MD, PhD)  
INSERM U823/ Team 11 (Outcome of Cancer and Critical Illnesses),  
Coordinating Center of the OUTCOMEREA Database  
Albert Bonniot Institute, La Tronche  
38706 CEDEX; France  
Tel : 33(0)4 76 76 87 79  
Fax : 33(0)4 76 76 55 69  
Email: jf.timsit@outcomerea.org.

### ABSTRACT

The sources of intensive care-related information and the means of communication increase rapidly. We presented here an overview of what should be done to collect high quality database. In a second part, the principle of the choice of the research question, the outcome, the

explanatory variables and the statistical methods to address the question are overviewed, emphasizing major and frequent pitfalls which should be avoided.

**Keywords:** Information systems; Database, factual; Biomedical research; Evaluation studies; Data collection/methods; Quality control

### INTRODUÇÃO

Nowadays, the need for accurate information usable not only in clinical practice and health-services but also for clinical audits and evaluation research is well recognized. Obviously, the proponents of each of these activities should combine their resources to obtain the data they need. Indeed, there is a growing recognition of the potential value of such databases not only for clinical audits but also to conduct clinical studies and evaluate health technologies.<sup>1</sup>

In order to achieve these objectives, clinicians, managers' consumers and researchers should define the variables that are needed, and propose complete and accurate definitions based on standard definitions of clinical disorders, interventions and outcome.

High quality clinical databases could help to provide up-to-date estimates of the probabilities of different outcomes in typical intensive care unit (ICU) setting. It should help to determine risk factors and prognostic models of various events and might be used for resource allocations between regions or hospitals.

The large number of patients included in databases enables sub-group analyses which, in turn, could provide clinicians with information on specific categories of patients. Similarly it could be used for studying rare disorders or interventions.

Research using a high quality clinical database is relatively cheap since the data-collection system body is established, their costs are spread over many research studies and is shared with the other applications such as the clinical management, audit and administrative uses.

Research on clinical database should not replace randomized control trials but have, in our opinion, a major role in demonstrating evidences.<sup>2</sup>.

First, they could generate hypothesis about risk factors and way of preventing a particular disease. Second, by raising the level of uncertainty among clinicians they might increase clinician likelihood of participating in a randomized control trial (RCTs). Third, they provide a permanent infrastructure for designing large multicenter trails. Fourth, it could be used to confirm on a general unselected population of patients the conclusions of RCTs, most often performed in highly selected specialized teaching ICUs.

In this manuscript we will describe how to construct a high quality database in order to use it for research purposes and to scheme what are the main way for exploring ICU events and prognosis.

### **HOW TO OBTAIN A HIGH QUALITY DATABASE?**

Outcome is measured by mortality, morbidity (iatrogenic events, nosocomial infections) functional status, quality of life, cost of care, length of stay, time to return to work, patient and family satisfactions. Factors that can influence the quality of the outcome data are the methods by which the data are collected, standardization of definitions, the currentness of the database, adequate number of patients and outcome.

There are numerous shortcomings of large databases including not capturing the unusual risk factors, and thus underestimating the risk of a given patient, lack of homogeneity of varying patient populations, and the accuracy of the data acquisition.

We proposed some important rules<sup>3</sup> used in well recognized ICU databases<sup>4,7</sup>.

#### *- At the onset of the database*

The data must be defined precisely and data characteristics must be in a data dictionary. A minimal data set should be set up.

The data-collection protocol should be written down, featuring a definition of pitfalls in data collection and creation of manual or automated data checking.

Ultimately, the central coordinating center should create a user-friendly case-record-form preferably using an electronic format.

Ideally, because the cohort must represent the overall ICU population, every patient must be entered in the database.

ICUs accepting to participate to the data entry must be aware of a quality insurance plan and must accept some rules:

New participant sites must be visited and training of

the new participants performed.<sup>8</sup> For new participants and periodically for others, data monitoring visits and quality audits must be realized using the case records or external data sources.

During the import of data into the central database, automatic checks must be performed.

#### *Communication with the centers*

It is important to develop a consistent and uniform communication with the practitioners involved in the data collection.<sup>9</sup> The data is owned collectively and it is important to allow opportunity to introduce new questions or ideas. It is also important to recognize the practitioner work (acknowledgement, authorship) and to organize data feedback. Publication rules concerning authorship should be defined early in the database creation and accepted by any new participating ICUs.

#### *- Quality improvement program*

Data quality reports are provided to centers. The centers should check detected errors, correct inaccurate data and fill in the incomplete data. After data audits or variability test, it is important to give feedback and recommendation to the center to resolve cause of data error.

Even for well-trained participants, a training course must be organized periodically in order to compute the same case-records and to unmask new causes of discrepancies. The cause and the consequences of errors should be discussed and will lead to an increase in the accuracy of the definitions. Interestingly, the motivation of the participants increases after the courses.

The central coordinating center is in charge of the adjustments of forms, software and dictionary, protocol and training material in order to include new variables or solve possible errors or discrepancies.

### **HOW TO CHOOSE THE STUDY DESIGN?**

The high quality database could be used to perform cohort studies. It could also be the core of nested case-control or exposed-unexposed studies. Cohort study is the most suitable design for conducting study in ICU<sup>10</sup>, because the follow up of patient is easy to collect as the large majority of events and risk factors occur during the hospital stay. The systematic inclusion of every patient entering the ICU ward favors the computation of outcomes prevalence.

Case Control studies are an alternative to cohort design comparing patients with an event (for example death) and others without; but they are rarely conducted

in ICU.<sup>11-12</sup> The matching of a case with several controls is possible to prevent from confounding factors. Case-control studies are usually faster and less expensive to conduct than cohort studies, but are very sensitive to bias (such as recall bias and selection bias).<sup>10</sup>

### AIM OF STUDY

The objectives of studies based on ICU databases are mostly associated to the research of risk factors of death or adverse events in the ward. In rarer cases, studies conducted in ICU are aimed to predict time dependant events in ICU<sup>13</sup> (i.e. number of infections during a year) or to model duration (e.g. length of stay).<sup>14</sup>

The first approach is to evaluate the feasibility of a study by ensuring that the frequency of the event of interest is large enough. The calculation of the statistical power is then needed. It consists in quantifying the ability to detect an association of any risk factors with the event of interest.<sup>15</sup> This figure can be computed using specific softwares.

### OUTCOME VARIABLE

The most commonly encountered qualitative variables in clinical study are binary outcomes (i.e. Dead or Alive). Early survival (Day 14, Day 28) mostly reflects the effect of the acute disease, but late survival (3-month, one year) more accurately reflect the impact of the disease for the patient and the family.<sup>14</sup> Time-variable end-point such as ICU survival or hospital survival should be discouraged as it should lead to biases in estimates.

Survival can be analyzed in its original form (by logistic regression) or completed with a variable describing the time-to-event (survival analysis). The first one is extensible for case-control design but does not consider exposure time, contrary to the other method that allows to control for the elapsed time before the occurrence (or not) of the event. In survival analysis, patients are censored if they do not undergo the event until their quit the study. Also, standard models assume the independence between censor time and event time (non informative censor). However, this strong assumption is frequently violated, particularly in ICU studies where the time to ward discharge and the time to death are totally dependant.<sup>17,18</sup>

In the case of a quantitative interest variable, it is advised to simply plot a histogram to make assumptions on the overall distribution pattern (Figure 1). Classical biological measures are often distributed according to

the Gaussian (Normal) distribution (e.g. age, Simplified Acute Physiology Score (SAPS II)) which is assumed by many of the statistical tests and allows the use of standard descriptive statistics (mean and standard deviation). But in some cases, variables are distributed otherwise (e.g. length of stay or the number of adverse events). Mean becomes really different from median. So first and last quartiles are preferred to mean and standard deviation which become inappropriate for data description with an absurd confidence interval (negative values). Otherwise, variable transformation can be applied to come closer to normality (e.g. log transformation). Atypical distributions and specially "multi peaks" pattern suggest cutting the variable in a number of classes equal to the number of peaks.<sup>19</sup>

In particular cases, the interest variable can be a subjective measurement of a phenomenon (e.g. Mac Cabe score, life quality score).<sup>20</sup> It is wiser to perform several measurements per subjects to insure for their reproducibility, thus reliability of the analysis.<sup>21,22</sup>

### EXPLANATORY VARIABLES

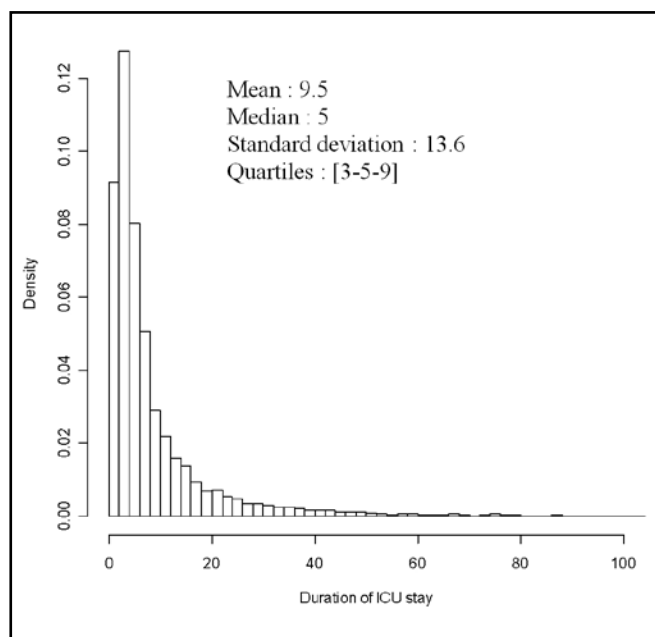
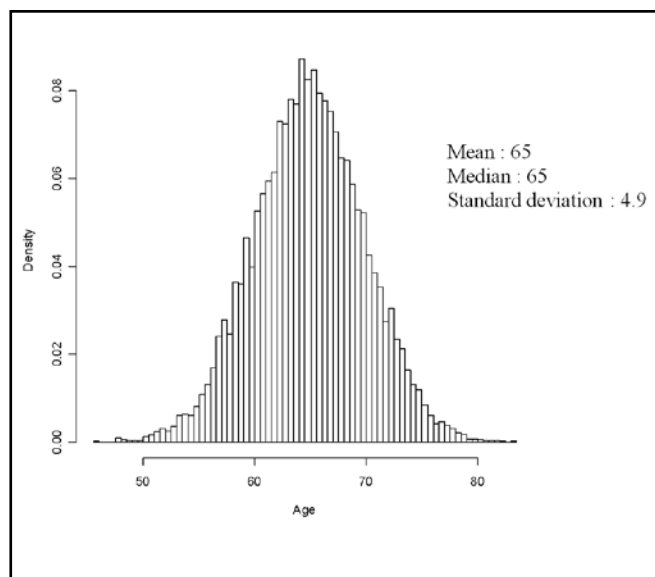
The description of the cohort data is an essential work which allows to verify consistency of data, missing value frequencies and to reveal the first interesting tracks to be investigated.

Data description techniques vary according to the possible variables type: quantitative (e. g., severity score like SAPS), qualitative binary (e. g., sex) or nominal (e.g., main symptom in ICU) and ordinal variables (e.g., American Society of Anesthesiology (ASA) codes). This can be done by contingency tables (frequencies and percentages) for categorical variable or mean, median, standard deviation, quartiles and histograms for continuous one.

Particular cautions should be taken with the counting of missing data to identify variables that are poorly reported. The multiplicity of the missing values in the data involves the deletion of observations in standard statistical analysis. Beyond 5% of missing data, it is wiser to drop the variable out of the study.<sup>10</sup> However, some efficient methods of data completion exist to deal with this issue.<sup>23</sup>

### RELATIONSHIP BETWEEN OUTCOME VARIABLE AND EXPLANATORY VARIABLES

Investigating for associations between explanatory variables and outcome variable is needed to highlight



**Figure 1 – Distributions of quantitative variables.** The histogram of age seems normally distributed: The mean and median are identical and correspond to the symmetric axis of the distribution. In this case, both mean  $\pm$  standard deviation or median and quartiles can be used to correctly describe data and confidence intervals can be easily computed. At the opposite, the distribution of ICU stay duration is asymmetric and even if the majority of ICU stays are short, extreme values can be identified (e.g. 90 days). In this case, the mean which is strongly influenced by extreme values is inappropriate to describe length of stay. Then, median and quartiles should be preferred to describe this variable

risk factors of the outcome event. This methodology named “univariate analysis” is described on Chart 1. The choice between parametric and non parametric tests is possible and assume (or not) an underlying probability law. The use of parametric tests is constantly more powerful but not always applicable (Chart 1). Variables listed from the literature, experience and common sense must be checked.

Assessing the effect of a quantitative variable on a binary outcome usually assumes that the increase in risk of the interest event is constant whatever its level. For example, the increase in risk of 28-day mortality is obviously lower between 15-25 years old than between 75-85 years old. This assumption of linearity of the risk increase must be checked by statistical tests (spline functions, log hazard).<sup>19</sup> If the test of this assumption is rejected for such variable, it should be recoded in several classes (e. g. quartiles or risk cut-offs). Otherwise, prefer the continuous one to keep the whole information.

### THE MAGICAL P-VALUE

But which p-value threshold do we have to choose? The statistical significance is based on the somewhat arbitrary choice of level (often 5 %). This psychological threshold of 5%, also named 5% Type I error (risk to conclude to a significant difference between two groups while that is caused by random) is not always relevant. For example, p-values of 0.048 and 0.052 are very close but conclusions are totally different. In the first case, we conclude to a significant difference between two groups and in the second, to the absence of difference between groups. So for each critical case, it is worth to think about medical interest rather only take into account statistical significance.

When lots of variables are tested, the risk to conclude to a significant association due to random increases with the multiplication of the tests performed (i.e. multiple comparisons). The commonly used error risk (p-value) is 5 % for one test, but if 4 tests are performed, this risk reaches 18.5%. Consequently, lower p values thresholds must be considered. This can be done using appropriate methods such as Bonferroni correction.<sup>19,24</sup>

### HOW TO CHOOSE AN APPROPRIATE MODEL?

The large panel of statistics tools existing allows the analysis of any type of variables and any study design. In medical research, the set of statistical models commonly used is restrained to the simplest ones because they do

not require too important skill in statistics to be understood and are considered robust.

Physicians must know the condition of validity of models and should be able to discuss with statisticians, however, at this step the help of experienced epidemiologist or statisticians is recommended.

Consequently, we present a selection of models that represent the extreme majority of what is used to analyze ICU data<sup>25</sup> in Chart 2.

Other models (or extensions) exist in addition to the previously ones mentioned, corresponding to more complex or less common study designs. Multilevel modeling allows taking into account the repetition of measures in a same subject or a hierarchical structure in the data (a patient in ICU ward, many ICU in a country, many countries in the world...)<sup>26-27</sup>. Forecasting and time series consist in analyzing the evolution of a series of correlated values in time (e.g. ARIMA models).<sup>28</sup>

**Chart 1 - Methodology of variables comparisons**

Type of interest variable	Type of variable tested	Methodology	Statistic test	Main assumptions
Qualitative (2 or more groups)	Quantitative	Descriptive statistics for each group (mean, standard error, median, quartiles...)	2 groups : T test * Wilcoxon k groups: ANOVA * Kruskall-Wallis	For ANOVA, dependent variables should be normally distributed within groups and variances are identical.
	Qualitative	Two-way table (cross-tabulation) of the frequencies and percentages	Chi-square Fisher exact	For Chi-square test, expected frequencies in each cell of table must be greater than 5.
Quantitative	Quantitative	Correlation coefficient	Pearson correlation* Spearman correlation	They identify linear relationship between interest and explanatory variable.
	Qualitative	Identical to the first part of this table (relation between qualitative and quantitative variables)		

\*Parametric test. ANOVA – analysis of variance

**Chart 2 - Different statistical models**

Type of interest variable	Statistical model to use	Main statistical assumptions to be verified	Other possible methods (non exhaustive)
Binary (e.g. death vs. alive; ill vs. healthy)	Logistic regression	There is a log-linear relationship between the independent variables and the odds of response. Predictive variables are independent one to another.	
Qualitative unstructured (e.g. diagnosis)	Polytomic logistic regression	There is a log-linear relationship between the independent variables and the odds of response. Predictive variables are independent.	Decision trees Discriminant analysis
Qualitative structured (e.g. life quality score ADL)	Poisson regression	Mean of response variable equal to the variance. Predictive variables are independent.	Negative Binomial regression Ordinal polytomic logistic regression
Continuous (e.g. rate of platelets in blood)	Linear regression	Linear relationship between explanatory variables and response variable. The residuals (predicted minus observed values) are distributed normally. Predictive variables are independent.	
Censored data (e.g. occurrence of ICU death)	Cox proportional hazard model	Proportional hazard assumptions. There is a log-linear relationship between the independent variables and the underlying hazard function. The censor should be non-informative (see text). Event frequency must be >5%	

ADL – activities of daily living, ICU – intensive care unit

Extensive survival models can integrate time dependent covariates or consider informative censoring.<sup>29,30</sup>

## VARIABLES SELECTION IN MODEL BUILDING

In the univariate analysis, we highlight all the associations existing with the outcome variable, without considering whether some explanatory variables were related one to each others. At the opposite, the multivariate analysis assumes independence of predictors, so one must not enter variables containing similar information (e.g. two organ dysfunction scores such as Logistic Organ Dysfunction score (LODS) and Sequential Organ Failure Assessment (SOFA))<sup>19</sup>. A few selection algorithms exist (backward, forward and stepwise) to help choosing variables to be included in the model. They automatically select a best subset of variables which are independently associated to the interest variable. These procedures are implemented in most of the statistical analysis softwares (SAS, Stata, SPSS, R...). However, it is largely preferable to select the one entered in the model based on medical and not necessary statistical rules such as the previously published literature, the easiness and reproducibility of data collection and the absence of missing data.

The model building will differ according to its purpose. A predictive model requires the inclusion of numerous explanatory variables to increase prediction ability (usually from 5 to 10% p-value thresholds). At the opposite, models aiming to search for independent risk factors need a more stringent selection of predictors (lower p-value for inclusion and stay in the model: usually from 1 to 5% thresholds). Indeed, variables can be forced in a model if they are deemed to be essential to the model interpretation (e.g. adjustment of the number of adverse events on the length of stay in a logistic regression). Great caution should be exerted when entering variables with rare modalities or subjective information. It increases variability and leads to unreliable estimates of the associated risk.<sup>31</sup>

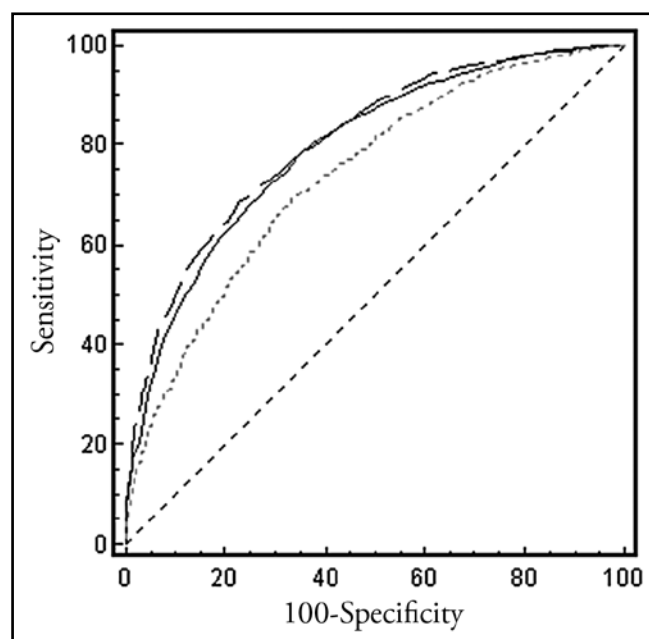
At last, introduction of clinically motivated interaction terms should be tested (like interaction between treatment and the severity of disease being treated) by adding cross-product terms.<sup>19</sup> Interaction terms increase rapidly if the number of variables entered into the model increases. For 3 variables, there is  $2^3-3=5$  interaction terms, for  $n$  variables there are  $2^n-n$  interaction terms. Therefore, it is recommended to test only clinically plausible and understandable (2-way) interaction terms.

Finally, there is no universal best model. So the choice of the final model should be motivated by the ease of use for clinicians (data quickly available in the unit, simplicity in calculation and interpretation) and to the ability of the model to fit the data.

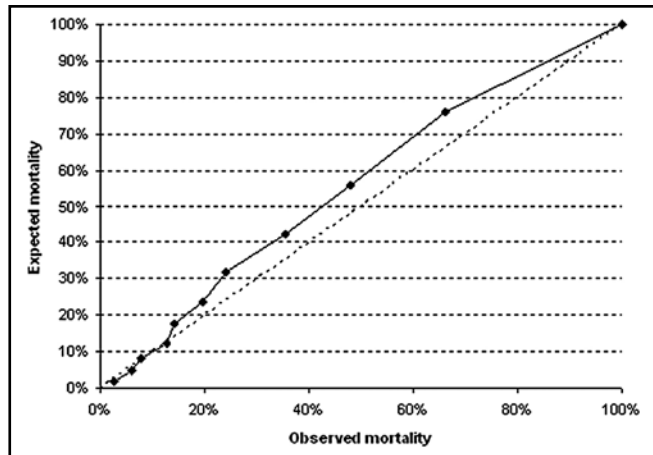
## MODEL FITTING AND VALIDATION

Model validation is essential part of the model development process, it is aimed to confirm its efficiency thus its utility in decision process. Standard prognostic model quality can be assessed thanks to discrimination and calibration.<sup>19,32</sup>

In a binary outcome case, the area under a receiver operating characteristics (AUC-ROC) curve estimates the sensitivity and specificity of the model to predict the outcome whatever the positivity threshold is. A value of 0.5 indicates no predictive discrimination (i.e. random) and a value of 1 indicates perfect separation of patients with different outcomes. A discrimination is correct from 0.80.<sup>19</sup> An example of ROC curve is described in Figure 2.



**Figure 2 – Receiver operating curves (ROC) example for three severity scores (MPM II 0, MPM III 0, SAPS 3) to mortality outcome, from OUTCOMEREA study.** The three ROCs represent the plot of sensitivity Vs 100- specificity while varying the scores discrimination threshold value. The area under the curve (AUC) represents the discrimination ability of the severity scores which is compared to the area under the diagonal lines (discrimination ability of the random = 0.5)



**Figure 3 – Calibration of the SAPS 3 score (from OUTCOMEREA study).** The calibration graph represents predicted mortality rate stratified by 10% intervals of mortality risk against observed mortality rate. In a well calibrated model, expected mortality within group should correspond to the observed mortality (diagonal line). The test of this hypothesis can be done by the Hosmer and Lemeshow test

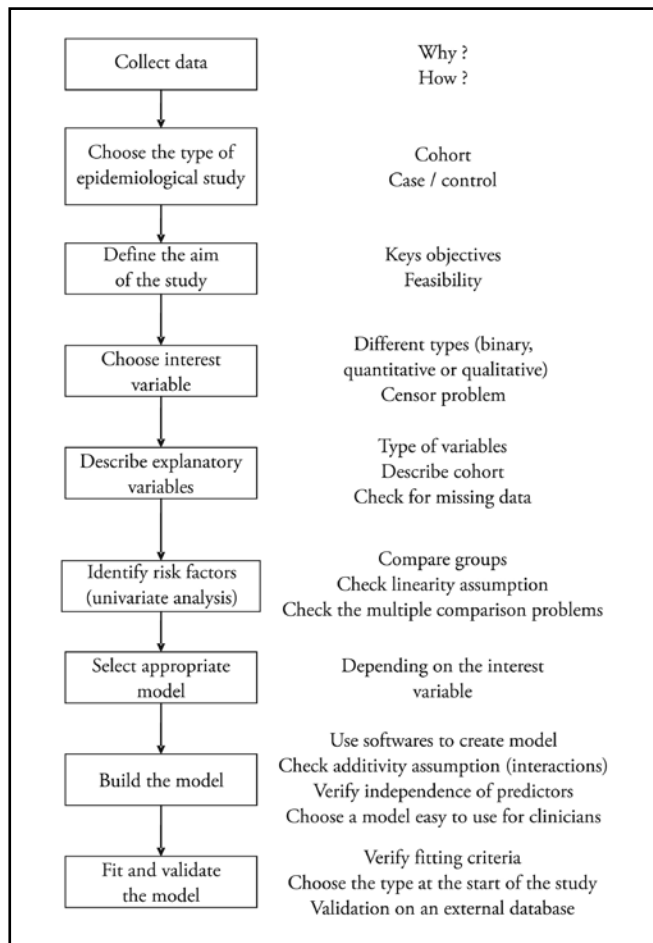
Calibration (or fit) component describes globally the predictive accuracy by checking bias between predicted and observed responses (see details of Hosmer-Lemeshow Chi2 test).<sup>32</sup> For example, if the average predicted mortality for a group of similar patients is 0.3 and the actual proportion dying is 0.3, the predictions are well calibrated. A graph can be drawn like in Figure 3.

Validation of a prognostic model can be done using the same cohort (internal: data-splitting, bootstrap and cross validation) or using patients from another cohort (external). This last method is the most reliable but more difficult to realize in practice. Details on these techniques are available in the literature<sup>19,33</sup> and this choice must be done at the start of the study.

### CONCLUSION

A flowchart summarizing the different steps that were developed previously in this document and that should be followed to conduct a multivariate analysis in ICU can be found in Figure 4.

Nowadays, more and more data, statistical tools, journals or scientific information through Internet are available. To make appropriate conclusion from a study, one must cautiously decide the precise methodology to be used before doing any data analysis in order to decrease the risk to draw a conclusion due to random or biases. Collaboration between ICU physicians and epidemiologists and/or biostatisticians is mandatory to avoid excessive comparisons and numerous pitfalls.



**Figure 4: Steps to follow to conduct a multivariate analysis.**

### RESUMO

As fontes de informação relacionadas a terapia intensiva e os meios de comunicá-la aumentam rapidamente. Neste artigo, nós apresentamos um panorama geral sobre o que deve ser feito para se obter dados de alta qualidade nas unidades de terapia intensiva. Os princípios para a escolha da questão clínica da pesquisa, os desfechos, as variáveis explanatórias e os métodos estatísticos a serem utilizados para abordar a questão são comentados de forma geral, com ênfase nos erros e ciladas mais freqüentes que devemos evitar.

**Descritores:** Sistemas de informação; Base de dados factuais; Pesquisa biomédica; Estudos de avaliação; Coleta de dados/métodos; Controle de qualidade

## REFERENCES

01. Black N. High-quality clinical databases: breaking down barriers. *Lancet*. 1999; 353(9160):1205-6. Comment in: *Lancet*. 1999; 354(9172):75.
02. McKee M, Britton A, Black N, McPherson K, Sanderson C, Bain C. Methods in health services research. Interpreting the evidence: choosing between randomised and non-randomised studies. *BMJ*. 1999; 319(7205):312-5.
03. Arts DG, De Keizer NF, Scheffer GJ. Defining and improving data quality in medical registries: a literature review, case study, and generic framework. *J Am Med Inform Assoc*. 2002; 9(6): 600-11. Review.
04. Aegerter P, Auvert B, Buonamico G, Sznajder M, Beauchet A, Guidet B, et al. [Organization and quality control of a clinical database on intensive care medicine in central and suburban Paris]. *Rev Epidemiol Sante Publique*. 1998; 46(3): 226-37. French.
05. Harrison DA, Brady AR, Rowan K. Case mix, outcome and length of stay for admissions to adult, general critical care units in England, Wales and Northern Ireland: the Intensive Care National Audit & Research Centre Case Mix Programme Database. *Crit Care*. 2004; 8(2): R99-111.
06. Stow PJ, Hart GK, Higlett T, George C, Herkes R, McWilliam D, Bellomo R; for the ANZICS Database Management Committee. Development and implementation of a high-quality clinical database: the Australian and New Zealand Intensive Care Society Adult Patient Database. *J Crit Care*. 2006; 21(2): 133-41.
07. Timsit JF, Fosse JP, Troché G, De Lassece A, Alberti C, Garrouste-Orgeas M, Bornstain C, Adrie C, Cheval C, Chevret S; For the OUTCOMEREA Study Group, France. Calibration and discrimination by daily logistic organ dysfunction scoring comparatively with daily sequential organ failure assessment scoring for predicting hospital mortality in critically ill patients. *Crit Care Med*. 2002; 30(9): 2003-13. Comment in: *Crit Care Med*. 2002; 30(9):2151.
08. Arts DG, Bosman RJ, de Jonge E, Joore JC, de Keizer NF. Training in data definitions improves quality of intensive care data. *Crit Care*. 2003; 7(2):179-84.
09. Proctor SJ, Taylor PR. A practical guide to continuous population-based data collection (PACE): a process facilitating uniformity of care and research into practice. *QJM*. 2000; 93(2):67-73.
10. Bowers D. *Medical Statistics from Scratch*. New York: John Wiley & Sons; 2003.
11. Adrie C, Azoulay E, Francois A, Clec'h C, Darques L, Schwebel C, Nakache D, Jamali S, Goldgran-Toledano D, Garrouste-Orgeas M, Timsit JF; OutcomeRea Study Group. Influence of gender on the outcome of severe sepsis: a reappraisal. *Chest*. 2007; 132(6):1786-93.
12. de Lassece A, Timsit JF, Tafflet M, Azoulay E, Jamali S, Vincent F, Cohen Y, Garrouste-Orgeas M, Alberti C, Dreyfuss D; OUTCOMEREA Study Group. Pneumothorax in the intensive care unit: incidence, risk factors, and outcome. *Anesthesiology*. 2006; 104(1):5-13.
13. Moine P, Timsit JF, De Lassece A, Troché G, Fosse JP, Alberti C, Cohen Y; OUTCOMEREA Study Group. Mortality associated with late-onset pneumonia in the intensive care unit: results of a multi-center cohort study. *Intensive Care Med*. 2002; 28(2):154-63.
14. Perez A, Chan W, Dennis RJ. Predicting the length of stay of patients admitted for intensive care using a first step analysis. *Health Serv Outcomes Res Methodol*. 2007; 6(3-4):127-38.
15. Machin D, Campbell MJ, Fayers PM, Pinol APY. *Sample size tables for clinical studies*. 2a ed. Oxford: Blackwell Science; 1997.
16. Angus DC, Laterre PF, Helterbrand J, Ely EW, Ball DE, Garg R, Weissfeld LA, Bernard GR; PROWESS Investigators. The effect of drotrecogin alfa (activated) on long-term survival after severe sepsis. *Crit Care Med*. 2004; 32(11):2199-206. Comment in: *Crit Care Med*. 2004; 32(11):2347. *Crit Care Med*. 2004; 32(11):2348-9.
17. Chevret S. Logistic or Cox model to identify risk factors of nosocomial infection: still a controversial issue. *Intensive Care Med*. 2001; 27(10):1559-60. Comment on: *Intensive Care Med*. 2001; 27(8):1254-62.
18. Cox D. Regression models and life-tables. *J Royal Stat Soc*. 1972; 34(2):187-220.
19. Harrell FE, Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996; 15(4):361-87.
20. Afessa B, Keegan MT. Predicting mortality in intensive care unit survivors using a subjective scoring system. *Crit Care*. 2007; 11(1): 109. Comment on: *Crit Care*. 2006;10(6):R179.
21. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960; 20(1):37-46.
22. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979; 86(2):420-8.
23. Molenberghs G, Kenward M. *Missing data in clinical studies*. New York: Wiley InterScience; 2007.
24. Westfall PH, Tobias RD. *Multiple comparisons and multiple tests using SAS system*. SAS Publishing; 2003.
25. Katz MH. *Multivariable analysis: a practical guide for clinicians*. 2nd ed. Cambridge; New York: Cambridge University Press; 2006.



26. Twisk JWR. *Applied multilevel analysis: a practical guide*. New York: Cambridge University Press; 2006.
27. Goldstein H, Browne W, Rasbash J. Multilevel modelling in medical data. *Stat Med*. 2002; 21(21):3291-315.
28. Charbonneau P, Parienti JJ, Thibon P, Ramakers M, Daubin C, du Cheyron D, Lebouvier G, Le Coutour X, Lelercq R; French Fluoroquinolone Free (3F) Study Group. Fluoroquinolone use and methicillin-resistant *Staphylococcus aureus* isolation rates in hospitalized patients: a quasi experimental study. *Clin Infect Dis*. 2006; 42(6):778-84. Comment in: *Clin Infect Dis*. 2006; 42(6):785-7.
29. Kim HT. Cumulative incidence in competing risks data and competing risks regression analysis. *Clin Cancer Res*. 2007;13(2 Pt 1): 559-65.
30. Resche-Rigon M, Azoulay E, Chevret S. Evaluating mortality in intensive care units: contribution of competing risks analyses. *Crit Care*. 2005; 10(1):R5. Comment in: *Crit Care*. 2006; 10(1):103.
31. Spiegelman D, Carroll RJ, Kipnis V. Efficient regression calibration for logistic regression in main study/internal validation study designs with an imperfect reference instrument. *Stat Med*. 2001; 20(1): 139-60.
32. Hosmer DW, Lemeshow S. *Applied logistic regression*. New York: Wiley; 1989.
33. Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med*. 2000; 19(4):453-73.